DOCUMENT RESUME

ED 377 278 UD 030 206

AUTHOR Jones, Russell W.
TITLE A New Model for School Accountability: The Urban District Assessment Consortium.
INSTITUTION Boston Coll., Chestnut Hill, MA. Center for the Study of Testing, Evaluation, and Educational Policy.
PUB DATE Aug 94
NOTE 37p.; Paper presented at the Annual Meeting of the American Psychological Association (102nd, Los Angeles, CA, August 1994).
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Accountability; Cooperation; Culture Fair Tests; *Educational Assessment; Elementary Secondary Education; Mathematics Instruction; Models; Reading Instruction; *School Districts; Standardized Tests; Test Bias; Test Construction; Test Items; *Urban Schools; Writing (Composition)
IDENTIFIERS Authentic Assessment; Open Ended Questions; *Performance Based Evaluation

ABSTRACT

This document contains a brief overview and introduction to the Urban District Assessment Consortium (UDAC), sample performance items with student responses, and sample open-ended items with student responses. The UDAC is a multiyear cooperative effort that is comprised of 11 major urban school districts across the United States and the Center for the Study of Testing, Evaluation, and Educational Policy at Boston College (Massachusetts). The goal of the UDAC is to develop and disseminate authentic-assessment instruments as alternatives to the traditional standardized tests for school-accountability purposes. The UDAC is based on the belief that the relation between classroom instruction and accountability assessment should be an integrative one. The UDAC's instruments, which include open-ended and performance-based tests in reading, writing, mathematics, and science for grades 4 and 8, have been developed to reflect emerging national standards in each of these areas. The UDAC assessments are racially, ethnically, and linguistically inclusive, and are offered in English and Spanish to accommodate the language needs of the greatest number of urban students. Appendix 1 contains sample performance-based test items and student responses, and Appendix 2 contains sample open-ended test items and student responses. (Contains 3 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 377 278

A New Model for School Accountability:

The Urban District Assessment Consortium

Russell W. Jones

Center for the Study of Testing, Evaluation, and Educational Policy
Boston College
323 Campion Hall
Chestnut Hill
MA 02167
Phone (617) 552-4536, Fax (617) 552-8419
Internet jonesru@hermes.bc.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R.W. Jones.
Boston
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CG030206

August 12, 1994

Paper presented at the conference of the American Psychological Association, Los Angeles, California, August 1994.

BEST COPY AVAILABLE

A New Model for School Accountability:

The Urban District Assessment Consortium

Russell W. Jones
Boston College

This document contains a brief overview and introduction to the Urban District Assessment Consortium (UDAC), sample performance items and sample student responses, and sample open-ended items and sample student responses. Additional information may be obtained by contacting either UDAC, or the Center for the Study of Testing, Evaluation, and Educational Policy at Boston College.

Overview

- The Urban District Assessment Consortium (UDAC) is a multi-year cooperative effort funded by grants from the John D. and Catherine T. MacArthur, Ford and Boston Foundations and Pew Charitable Trusts. The consortium is comprised of 11 major urban school districts across the United States and the Center for the Study of Testing, Evaluation, and Educational Policy at Boston College.

- Established two and a half years ago the goal of UDAC has been to develop and disseminate authentic assessments as alternatives to traditional standardized tests for school accountability purposes. While its primary purpose is to serve school level accountability needs, UDAC is based on the belief that the relationship between classroom instruction and accountability assessment should be an integrative one. UDAC's instruments, which include open-ended and performance tests in reading, writing, math and science for grades 4 and 8, have been developed to reflect emerging national standards in each of these four subject areas.

- While the need for new ways of assessing student learning has been well documented, the vast majority of research conducted in this area has been limited to classroom assessments, such as student portfolios. There remains a dearth of knowledge regarding the use of these new assessments as accountability instruments.

- Much of UDAC's research and development has focused on addressing the needs of the "at risk" urban student population. Previous reform efforts aimed at bolstering the performance of these students have met with mixed results, and past reform measures focusing on assessment, such as the minimum competency and the high stakes testing movements, have been shown to be particularly damaging to this population of students (Lomax, West, Harmon, Viator & Madaus, 1992; Madaus, 1991; Shepard & Dougherty, 1991). With a commitment to improving the educational opportunities of urban students, and a firm belief in the power of expectations to influence performance outcomes, UDAC assessments articulate and demand rigorous performance standards for *all* students. In addition, UDAC assessments are racially, ethnically and linguistically inclusive, and are offered in English and Spanish to accommodate the language needs of the greatest number of urban students.

The CSTEEP - School Community Partnership

Staff at the Center for the Study of Testing, Evaluation, and Educational Policy (CSTEEP) at Boston College assist schools in:

1. Identifying educational goals and assessment strategies.
2. Designing and conducting training for school and district staff in the:
    i. Use of assessment techniques.
    ii. Administration of assessment instruments.
    iii. Reporting of results to parents and the community.
3. Designing matrix sampling techniques for school-level accountability purposes.

Participating school districts:

1. Identify individuals to act as a liaison between the schools, the district and UDAC staff.
2. Accommodate the increased in-service demands of the participating schools.
3. Explore, in conjunction with UDAC staff, how to compare new assessment practices with current testing requirements.
4. Involve parents and the business community in school improvement efforts.
5. Will continue to implement the new assessment systems after the project is completed.

UDAC Instruments

The UDAC assessment administration reported here assessed the performance of 302 fourth grade students within six Boston Public Schools, and 209 eighth grade students within four Boston Public Schools. Performance was assessed using performance items, long and short response open-ended items, and NAEP (predominantly multiple-choice) items. This approach, using multiple item formats, provided teachers, students and other members of the school community with a wealth of information. In particular, it is worthwhile to highlight the large amount of information obtained from the performance items. Sample performance items and student responses for reading, writing, math and science assessments are provided in Appendix I. Sample open-ended items and student responses for reading, writing, math and science assessments are provided in Appendix II.

Teachers, especially, were impressed with the quantity of pedagogical information obtained whilst (1) observing students perform the assessment tasks and (2) when scoring student responses. Because of the close ties between UDAC and parents, business and other community members, the broader school community also benefited from the assessment administrations by getting a clear indication of student abilities and performance across a range of tasks.

Systematic matrix sampling was used during the assessment administration. This procedure permits the use of a wider range of assessment techniques than if all students were tested with the same assessment instrument. Similar judicious use of samples has long been recognized in accounting, business and industry when auditing or testing for quality control. For example, General Motors, Toyota or Ford only crash test a sample of the cars they build. Yet, these companies use the results of these few tests to draw conclusions concerning the performance of all cars they produce. Systematic matrix sampling was found to be extremely

BEST COPY AVAILABLE

effective in assessing student performance across the four subject areas (math, science, reading and writing) using a variety of assessment techniques.

To meet the needs of both English and Spanish speaking students within the target schools, items were developed in both English and Spanish. This bilingual approach to item development was an important aspect of UDAC assessments. UDAC did not merely translate items from English to Spanish. Instead items were developed simultaneously in the two languages. This permitted item writers to focus on content and vocabulary from the outset and to immediately address any differences in syntax, semantics or concepts. Using this approach items were not centered in a single language or culture. Special attention was paid to using words and idioms that spanned the several Latino cultures, rather than focusing on "textbook" Spanish or a single Latino culture. Item development teams were composed of native speakers of both languages. These teams brought to the item development process diverse perspectives extremely useful for the creation of truly multicultural assessment instruments. The approach adopted by UDAC minimized many of the linguistic problems encountered by many testing programs whereby all items are developed in a single language, and then translated into what is usually termed "the second language."

Scoring Procedures

Student responses to NAEP items were scored using NAEP answer keys. Student responses to performance and open-ended items were scored using a five point scoring rubric. A generalized rubric (shown on page 6) was developed to serve as an overall guide, although an item specific rubric was also developed for each item.

Generalized Rubric used in Rating UDAC

Open-Ended and Performance Items

4 - Excellent Response: Response is correct, complete and communicated effectively.

3 - Adequate Response: Response is essentially correct or clear despite minor errors or omissions.

2 - Inadequate Response: Response is poor, incorrect, or reflects a misconception or miscue.

1 - Too Brief: Too brief to evaluate, no response, item left blank, restates question, "I don't know."

0 - Student Absent.

This rubric applies uniformly across subject areas and grade levels. Sample answers are provided in the item specific scoring keys.

Rubrics were developed under the guidance of subject matter specialists. Item writing and rubric development occurred concurrently. Scoring was performed by teachers, parents, other community members, and UDAC staff. Prior to the scoring of student responses, training sessions were held to train personnel in scoring procedures and to provide practice in rating student responses.

Rater Teams

Items were scored by three member teams comprised of a UDAC representative, a school representative (usually a teacher or administrator), and a member selected from the community. Each performance and open-ended item was independently scored by two members of a rating team. Following this initial independent scoring, team members met and compared their scores. If discrepancies in scoring between two raters occurred, the two raters discussed their reasons for assigning a particular score in an attempt to reach consensus. Typically this discussion resulted in raters agreeing to alter one or both ratings

to reach consensus. Consensus was reached when the two raters agreed on a score. If consensus was not reached both scores were recorded.

Summary of Results

Analysis of UDAC performance and long and short response open-ended items showed most functioned very well. Item difficulty values for performance items covered a broad range from .33 to .69 in reading, from 0 to .69 in science, from 0 to 1.0 in math and from .26 to .47 in writing. Item difficulty values for open-ended items also covered a broad range from 0 to .73 in reading, from 0 to .87 in science, from 0 to .88 in math and from 0 to .83 in writing. Thus, for each subject, the assessment instruments were able to assess students with a wide range of ability levels from very weak to very strong. Item discrimination values for most items were generally well above the widely accepted value of 0.3. Indeed, the mean item discrimination value for UDAC performance items was .50 in reading, .35 in science, .47 in math, and .37 in writing. However, values for open-ended items were generally lower; .46 in reading, .27 in science, .23 in math, and .38 in writing. Thus, UDAC items provided a great deal of information and were able to differentiate between students having various levels of ability in each of the subject areas. Statistical analysis of item properties identified a small proportion of items that appeared to be functioning poorly. These items have been revised or eliminated prior to further test administrations.

In addition to the ability of UDAC items to assess and discriminate well between students of varying ability, an evaluation of the relationship between the Metropolitan Achievement Test (MET) (a standardized test given to students in the Boston Public Schools district) and NAEP standardized tests shows UDAC items provide different information to that provided by these standardized tests. All three assessments were administered to students during the same semester. The NAEP and MET assessments are both established testing tools widely accepted as suitable measurement tools to evaluate student ability (for example, NAEP

tests rank among the most established and broadly applied tests in the United States having been used to evaluate the national performance of American students for more than 20 years). Correlational analyses between these two assessments and UDAC assessments revealed a moderate (for example, r = .31) to high (for example, r = .66) positive relationship between student performance on the MET and NAEP tests and their performance on the UDAC assessments. This relationship provides powerful criterion-related validity for the newer UDAC assessments. However, the finding that this relationship is not perfect clearly indicates UDAC assessments provide a substantial amount of information different to that provided by either the MET or NAEP standardized tests.

School faculty and community members were very appreciative of the wealth of information provided by UDAC items. By participating in the administration and scoring of student responses they were presented with clear demonstrations of student ability. In other words, educational feedback was not limited to post analysis consideration of assessment results, instead the entire assessment process was a learning activity for all individuals who were associated with the assessments. Evaluation of the UDAC program included written comments by school faculty. These comments point to the benefits obtained as a result of the UDAC assessments. In the words of one principal "The greatest asset (of these assessments) is that there are implications for teaching!"

References

Lomax, R. G., West, M. M., Harmon, M. C., Viator, K. A., & Madaus, G. F. (1992). *The impact of testing on minority students*. The influence of testing on teaching math and science in grades 4-12, Topical Paper. Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Madaus, G. F. (1991, April). *The effects of important tests on students: Implications for a national examination or system of examinations*. Paper presented at the AERA Invitational Conference on Accountability as a State Reform Instrument: Impact on Teaching, Learning, Minority Issues and Incentives for Improvement, Washington, D.C.

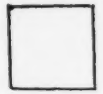
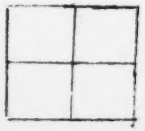
Shepard, L. A., & Dougherty, K. C. (1991, April). *Effects of high-stakes testing on instruction*. Paper presented at the annual joint meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago.

APPENDIX I

Sample Performance Items and Student Responses

MATHEMATICS	MIDDLE LEVEL

Directions: Read the directions carefully as you complete this activity. In your envelope you will find some small toothpicks. You need to use them to answer the questions below. You will make the shapes, draw a picture and count the number of toothpicks you used.

Make	Draw	# toothpicks used
EXAMPLE: The smallest number of toothpicks you need to make one square.		4
The smallest number of toothpicks you need to make two squares. [EMM20-P01] [EMM25-P01]		7
The smallest number of toothpicks you need to make three squares. [EMM25-P02] [EMM20-P02]		10
The smallest number of toothpicks you need to make **exactly** four squares. (Be careful. This is a tricky question.) [EMM25-P03] [EMM20-P03]		13
Draw two different ways to make five squares. Use no more than sixteen toothpicks. [EMM300-04]		16 14 12

MATHEMATICS	MIDDLE LEVEL

ACTIVITY #2

Directions: Use the graph paper provided to draw the following figures.

5. Draw a rectangle with a perimeter less than 24 inches. Label it "A." The perimeter of rectangle A is 20 inches.
[EMM30-P05]

6. Draw a rectangle with a perimeter one half that of rectangle A. Label it "B." The perimeter of rectangle B is 10 inches.
[EMM35-P06]

7. Draw a rectangle with an area twice that of rectangle A. Label it "C." The area of rectangle C is 48 square inches.
[EMM35-P07]

8. Draw a triangle with an area one half that of rectangle A. Label it "D." The area of triangle D is _____ square inches.
[EMM35-P08]

BEST COPY AVAILABLE

Directions: At this station you will find hot and cold water, several beakers, some Alka Seltzer tablets, and a thermometer. Plan an experiment to answer the question below.

How could you find out if the temperature of the water makes any difference in how fast a tablet will dissolve?

1. Describe below what you plan to do to find out what difference temperature makes on how fast a tablet dissolves. [EPS30-P05]

Pour hot water into a beaker and fill it up to the top. Put Alka Seltzer tablet into water and time how long tablet takes to dissolve. Do the same thing with cold water in another beaker.

2. Conduct the experiment and record all of your measurements in the space below. Be sure to put the correct units (milliliters or degrees Celsius, or degrees Fahrenheit, or minutes and seconds) beside each measurement. [EPS30-P06]

Kind of Measurement	Hot Water Beaker	Cold Water Beaker
Volume	50ml	50ml
Temperature	87°F	39°F
Time	41 seconds	58 seconds

3. What did you find out about the effect of water temperature on how fast the tablets dissolved? [EPS30-P07]

Tablets dissolve faster in warm water than cold water

BEST COPY AVAILABLE

4. If there was a difference in the speed of dissolving between the hot and cold, why do you think that happened?

[EPS30-P08]

The hot water and the Alka Seltzer reacted together than the cold water and the Alka Seltzer

5. List some ways your group thinks you might improve the experiment so that your results could be more accurate.

[EPS30-P09]

We should do the experiment several times.

6. Do you think you would get the same results if you had used a completely different substance such as salt or sugar or aspirin? Why or why not?

[EPS30-P10]

No. Because different substances are made of different things and so they behave differently in water.

BEST COPY AVAILABLE

Directions: All of the questions in this booklet are about the story A Bicycle for Rosaura. You need to do all of the activities in the booklet.

1. Think about the story A Bicycle for Rosaura. Then write the answers to the questions in the spaces below.

What is the title of the story?
A Bicycle for Rosaura

What is the little hen's name?
Rosaura

What is Señora Amelia's problem?
She needed a bike for Rosaura

What steps does she take to solve her problem?
She asked the man to make a bike

[EPR25-P1]

2. How does Rosaura feel when she rides her bicycle down the street every morning?

Rosaura might feel very happy.

[EPR25-P2]

5. In the envelope you will find five pictures. These pictures come from the story A Bicycle for Rosaura. Arrange the pictures in the order that they happened in the story. Once you have put the pictures in order, number them from 1 to 5 in the boxes with the # in them.

[EPR25-P5]

a- Put the pictures in order.
b- Number the pictures 1 to 5.

6. Using the tape recorder, retell the story A Bicycle for Rosaura in your own words.

[EPR25-P6]

BEST COPY AVAILABLE

This is a copy of a mural located on the Rafael Hernandez school in Roxbury, MA. Community artist Roberto Chao painted this mural in 1990 with the help of Egleston Square residents and some students from the school.

1a. Pick one thing in the mural that is familiar to you, that is important to you or that is part of your culture. Describe it in detail and explain why you chose it. Use a lot of details, description, and examples in your writing. Write as much as you can. [EMWINP-1]

I like the part where it has the guitar, maracas, and drums. The different kinds of music. The land with the palm trees. That's the part of the mural I like because it has caribbean stuff. I from the caribbean so I know about that. I like to learn different kinds of dance. I love merengue because its the national dance of Dominican Republic where I'm from. I also like salsa. The land where I see it as a painting or picture I want to be there. The music makes everybody move to the rhythm. The part I choose is cuz I like people of others.

BEST COPY AVAILABLE

APPENDIX II

Sample Open-Ended Items and Student Responses

1. Tomas' teacher ordered lunch for his class from a local sandwich shop. She ordered the lunches from the menu below.

Sandwiches	*Milkshakes*
ham & cheese	chocolate
roast beef	vanilla
turkey club	strawberry
bacon, lettuce & tomato	pineapple

If a lunch consists of one sandwich and one milkshake, how many different lunch combinations are possible? [EMM30A2-22]

4

Answer: 16

2. If the teacher ordered one of each of the possible lunch combinations, what is the probability that Tomas would receive a turkey club sandwich and a chocolate milkshake for lunch? [EMM30A2-23]

4

Answer: 1/16

4. Below is a third grader's addition homework. Look carefully at each answer. The student is very confused.

29	44
+ 3	+ 6
212	410
18	67
+ 67	+ 84
715	1411

On the lines below, write down everything you think this student did wrong and how you would help the student understand why the answers are wrong.

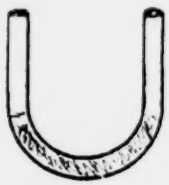
The student never placed the first digit (of all the 2-digit #'s) on top of the tens place #'s to add together, instead he wrote the full ~~#~~ number as part of the solution and then continued to add. If he just put the number at the top (instead of below the line) the answers would be correct.

[EPM30A2-25]

Directions: Read each question carefully before answering.

Jane poured some water into the U-shaped glass tube as shown here. However, there was not enough water to completely fill the tube.

3. Draw on the picture below what the water level in the tube will look like when all the water has been poured in. [EMS25A2-3]

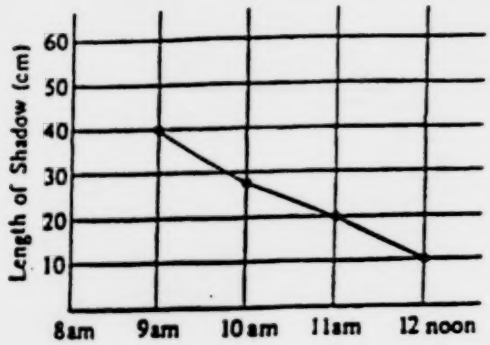


4. Why do you think it will look the way you have drawn it? [EMS30A2-4]

gravity pulls the water down it wouldn't have stoped on the long side of the tube

Directions: Think about each question carefully before writing your answer.

Kim placed a stick in the ground on a sunny day. Then she measured the length of the stick's shadow at four different times. She made a graph of the measurements, which is shown to the right.



5. About how long was the shadow of the stick at 11 a.m.? Circle the correct answer below.

A. 10 cm.

(B.) 20 cm. 4

C. 30 cm.

D. 40 cm. [EMS20A2-5]

6. Why did the shadow get shorter and shorter? [EMS30A2-6]

2 The sun moves and as it gets above the stick the shadow gets shorter

Directions: The following passage is taken from a book by Rigoberta Menchú. Read it carefully and answer the questions that follow.

Rigoberta Menchú is a woman from Guatemala. She wrote the book called I, Rigoberta Menchú An Indian Woman In Guatemala. In 1992, she won the Nobel Peace Prize. She received this award because she spent most of her life fighting the terrible injustices she and her people suffered. In the following paragraphs, she tells us a little about her life as a child and the kind of work she and her mother did in Guatemala.

special vocabulary: *finca: a farm*
tortillas: unleavened bread made from cornmeal
nixtamal: dough used to make tortillas

"An Eight-Year-Old Agricultural Worker"

'And that's when my consciousness was born'
-Rigoberta Menchú

I worked from when I was very small, but I didn't earn anything. I was really helping my mother because she always had to carry a baby, my little brother, on her back as she picked coffee. It made me very sad to see my mother's face covered in sweat as she tried to finish her work load, and I wanted to help her. But my work wasn't paid; it just contributed to my mother's work. I either picked coffee with her or looked after my little brother, so she could work faster. My brother was two at the time....

I remember that, sometimes, my mother's work was making food for forty workers. She ground maize, made *tortillas*, put the *nixtamal* on the fire and cooked beans for the workers' food. That's a difficult job in the *finca*. All the dough made in the morning has to be finished the same morning because it goes bad. My mother had to make the number of *tortillas* the workers would eat. She was very appreciated by the workers because the food she gave them was fresh. The food we ate was cooked by another woman who sometimes gave us things that had gone bad, or *tortillas* which were tough and beans which jumped when you tried to pick them up. In the *finca* the women who do the cooking don't know which people they will cook for. The overseer comes and says, 'This is your group ... this is what you give them to eat, these are the people you feed, you feed them at such and such a time ... so get to work.' So different women fed us. My mother liked to give the workers the food they deserved, even if it meant she didn't sleep all night. They came back tired from the fields and she wanted to see that they ate well, even though her own family was eating badly somewhere else.

I was five when she was doing this work and I looked after my little brother. I wasn't

earning yet. I used to watch my mother, who often had the food ready at five o'clock in the morning for the workers who started work early, and at eleven she had the food for the midday meal ready. At seven in the evening she had to run around again making food for her group. In between times, she worked picking coffee to supplement what she earned. Watching her made me feel useless and weak because I couldn't do anything to help her except look after my brother. That's when my conciousness was born. It's true. My mother didn't like the idea of me working, but I did. I wanted to work, more than anything to help her, both economically and physically....

1. What kind of work did Rigoberta do when she was a child?

C He had to watch his brother or help his mother pick coffee

[EMR20A2-14]

2. What is an agricultural worker?

C a person that workes with the land or farming

[EMR25A2-15]

3. Why was her mother's job so difficult?

B Sometimes she had to cook for forty workers and if she didn't cook it all it would go bad

[EMR25A2-16]

4. Why was it so important for Rigoberta to help her mother?

D it let her work

[EMR25A2-17]

Directions: Read the following questions and write your answers in the space provided.

1. When people buy something and it breaks or doesn't do what it is supposed to do, they often write a letter complaining to the manufacturer about the product. Pretend you bought a very expensive pair of sneakers that fell apart the second day you wore them. On the lines below, write a letter to the company that makes the sneakers. Let the company know:

   - what kind of sneakers you bought
   - what was wrong with them, and
   - what you want the company to do about it.

   Make sure your letter is a business letter that includes the name and address of the sneaker company, a date, a greeting, the body of the letter, a closing and a signature. If you need more space, use the next page.

[redacted]

May - 19 - 94

Mr. Owner

I have a problem with your Shoe. A week ago on the 13th I Bough A pair of your Hightop Fila kicking. Two days later on the 15th I wore them to School and the laces ripped and the Sole came apart

I would like a full refund and a pair of quality Sneakers.

A former Buyer,

[redacted]

[EMWPA2-9]

2. In a paragraph, explain why you think the letter you wrote will make the company respond to your complaint.

Most people wouldn't bother to write a letter. This probably makes them think they may try to sue us. and I added "former Buyer" that means they lost a costomer.

[EMWIA2-10]